주석 및 내용 기반 검색을 지원하는 동영상 정보 관리 시스템

전미경[†] · 김인홍[†] · 류시국[††] · 전용기[†††] · 강현석[†††]

요 약

일반적으로 동영상에 대한 정보 검색에는 주석 기반 검색이나 내용 기반 검색을 사용한다. 그러나, 주석 기반 검색은 사용자의 주관이 개입되어 일관성을 잃기 쉽고 내용 기반 검색은 동영상 데이터가 담고 있는 일반적인 의미 추출이 어렵다는 단점을 가지고 있다. 그래서, 본 논문에서는 이 두 검색 기법을 상호 보완하여 검색의 효율성과 정확성을 높일 수 있도록 하는 통합 동영상 데이터 모델(IVDM)을 제안한다. 이 모델은 동영상 데이터를 분석하여 계층적으로 구조화한다. 상위 수준에서는 주제별로 부여된 메타 정보로 주석 기반 검색을 지원하고, 하위 수준에서는 동영상 데이터에서 색깔, 모양, 움직임, 질감 등의 특징 데이터를 추출하여 내용 기반 검색을 지원한다. 그리고 이 IVDM의 타당성을 입증하기 위해 동영상의 대표격인 뉴스 동영상을 객체 지향 데이타베이스 스키마 형태로 설계하고, 이에 맞는 질의 유형 4가지와 그 처리 알고리즘도 제공한다.

A Video Information Management System for Supporting Caption- and Content-based Searches

Jun Mi-Gyung[†], Kim In-Hong[†], Rhyoo Shi-Kook[††], Jun Yong-Kee[†††] and Kang Hyun-Syug[†††]

ABSTRACT

Generally, either caption-based search method or content-based search method is used to retrieve video information. However, each search method has its limitations. Caption-based search is apt to lose consistency as for user's subjects, and content-based search is hard to extract general means. To enhance efficiency and correctness as for complementing each other, we propose the Integrated Video Data Model(IVDM) which integrates the two search methods, to device the model, we analyze video data and construct the structure of video information hierarchically. IVDM supports caption-based search as assigning meta-data by analyzing thematic-unit in the higher level, and also supports content-based search as extracting feature data by analyzing the content of video data in the lower level. We design Object-Oriented database schema of news video, based on the IVDM. And we provide 4-type of queries and query processing algorithm to retrieve news video information.

1. 서 론

최근 멀티미디어와 인터넷 기술의 발달로 단순 문자 형식만이 아닌 화상, 소리, 동영상 등의 멀티미디어 정보, 특히 동영상 정보의 관리에 대한 수요가 급증하고 있다. 이에 따라 동영상 정보를 효율적으로 저장하고 필요에 따라 빠르게 검색할 수 있는 시스템에 대한 연구가 많이 이루어지고 있다[1-3].

이러한 연구들은 주로 동영상 정보를 검색하기 위해 2가지 접근법중 하나를 사용한다. 즉, 내용 기반

본 연구는 경상대 부설 정보통신 연구센터의 연구비 지원에 의한 것임

† 경상대학교 전자계산학과
†† 경남정보대학 컴퓨터정보과
††† 경상대학교 컴퓨터과학과

검색[6,12]과 주석 기반 검색[1,3]을 선택하여 사용하는 경향이 있다.

내용 기반 검색은 동영상 데이터에서 특징 데이터 즉 모양, 색깔, 질감, 움직임 등을 자동 추출하여 이를 바탕으로 검색한다[5,10]. 따라서 사용자의 주관이 개입되지 않아 객관적인 검색이 가능하고, 특정 정의역에 대해서 좋은 검색 결과를 나타낸다. 그러나 동영상 데이터가 담고 있는 일반적인 의미를 추출하기 어렵다는 단점이 있다. 반면 주석 기반 검색은 동영상 데이터가 담고 있는 의미를 사람이 먼저 파악한 후 이를 자연어를 이용하여 표현하고, 이를 바탕으로 검색하는 방법이다[1,3]. 이 방법은 자동화된 방법으로 알아내기 힘든 동영상 데이터의 여러 가지 의미를 쉽게 모델링할 수 있게 하고 이를 검색에 이용할 수 있게 한다. 그러나 사용자의 관점에 따라 주석을 다르게 부여하거나 다르게 해석할 수 있어 동영상 데이터에 대한 일관성을 잃기 쉽다. 특히 포괄적인 내용의 주석이 아닌 아주 세분화된 부분까지 주석을 부여하고자할 때는 더욱 더 일관성을 유지하기가 어려워진다. 따라서 이 두 접근법이 갖는 각자의 장점을 함께 수용하고 단점은 상호 보완하여 사용한다면 효과적일 것이다.

본 연구에서는 방대한 양의 동영상 데이터를 효율적인 방법으로 저장하고 검색할 수 있도록 일반적인 동영상 데이터를 분석하여 주석 기반 검색과 내용 기반 검색을 함께 사용할 수 있는 통합 동영상 데이터 모델(IVDM)을 제시한다. 그리고, 이 모델을 바탕으로 동영상의 대표격이라 할 수 있는 뉴스 동영상을 예로 객체 지향 데이터베이스 스키마를 설계하여 IVDM의 타당성을 입증하며, 이에 맞는 질의 유형 4가지와 그 처리 과정을 간단한 알고리즘으로 표현하다.

논문의 구성은 다음과 같다. 2장에서는 동영상 검색 기법에 관련된 연구를 살펴보고, 3장에서는 주석 기반 검색과 내용 기반 검색을 동시에 가능하게 하는 통합 동영상 데이터 모델을 개발한다. 4장에서는 3장에서 제시한 통합 동영상 데이터 모델을 이용하여 뉴스 동영상의 객체 지향 데이터베이스 스키마를 설계하고, 이에 알맞은 질의 유형과 질의 처리 알고리즘을 설명하며, 결과를 보여준다. 마지막으로 5장에서는 결론과 향후 연구 과제를 논의한다.

2. 관련 연구

동영상 데이터를 검색하는 방법에는 크게 내용을 기반으로 하는 검색과 주석을 기반으로 하는 검색으로 나눌 수 있다.

우선 동영상에 텍스트 형태의 정보(주석)를 부여하여 검색에 이용하는 주석 기반 검색 연구를 알아본다.

CVIMS[1]은 허진용 등이 동영상 문서의 구조를 추출하여 동영상 스트림의 인덱싱에 대한 의미를 부여할 수 있는 일반적인 동영상 데이터 모델을 제시하고, 이 모델을 이용하여 MPEG-2로 압축된 동영상 정보를 데이터베이스에 저장하고 검색할 수 있도록 개발한 MPEG-2 압축 동영상 정보 관리 시스템이다. 이 시스템은 MPEG-2 파일에서 I-프레임들을 추출하고, 그 중에서 키 프레임(대표 이미지)을 선택한 후, 선택된 키 프레임에 대한 축소그림(thumbnail)과 주석 및 그림 설명 정보를 데이터베이스에 저장하여 검색에 이용하였다.

VIRON[3]은 김기욱과 김형주가 개발한 주석 기반 동영상 검색 시스템으로 사용자간의 관점 차이에서 비롯된 주석 공유의 난점을 극복하기 위해, 단일의 주석 시스템 구조를 속성 스키마, 개념적 주석 구조, 동영상 인덱스 스트림 구조로 분할하여 동영상 데이터 질의 시스템을 모델링하였다.

CVIMS와 VIRON은 가능한 객관적인 관점에서 주석을 부여하여 공유하도록 설계를 하였지만 여전히 주석 기반의 한계라 할 수 있는 사용자의 주관성을 완전히 배제하지는 못하고 있다.

이와 달리 사용자의 주관이 전혀 개입되지 않는 동영상 데이터의 움직임과 같은 시간 차원과 색깔, 모양, 질감과 같은 공간 차원의 특징을 추출해 내어 검색에 이용하는 내용 기반 검색에 관한 연구도 많이 이루어졌다.

먼저 내용 기반 검색 시스템에서 사용되는 질의 유형을 보면 키워드를 입력으로 사용하는 경우와 그림을 그릴 수 있는 툴(tool)의 도움을 받아 사용자가 직접 스케치한 그림이나 파일로 존재하는 정지 영상(화상, Image)을 입력으로 사용하는 경우가 있다.

키워드를 사용한 예로 A. Ono[7] 등은 장면 묘사어(Scene Description Keyword)를 사용하여 170여개의 경치 화상을 검색하는 시스템을 제안하였고, Y. Kim과 M. Shibata[8]는 동영상의 내용을 자연어를

인덱스 트리 형태로 표현하여 검색시 이용하였다.

QBIC[10]과 Chabot[9], VisualSEEK[6]에서는 화상의 종류 키워드, 키워드와 함께 텍스쳐, 스케치, 모양을 사용한 대규모의 데이터베이스 상에서 검색할 수 있도록 연구하였다. 이 중 QBIC[10]은 IBM Almaden 연구소에서 개발된 것으로 내용 기반 화상 검색과 제한된 범위에서 동영상 검색이 가능하도록 하였다. 이 시스템에서는 사용자가 화상의 색상이나 질감, 모양 등과 같은 시각적 질의를 제공하거나 사용자가 찾고자하는 화상의 외곽선을 대략적으로 그리면 그것과 유사한 화상을 찾아준다. 또한 화상에 대해 적절한 키워드로 주석을 부여할 수 있다.

Chabot[9]은 Berkeley 대학에서 개발한 관계형 데이터베이스 시스템(INGRES)을 이용한 내용 기반 화상 검색 시스템으로, 화상에 대한 히스토그램(histogram)을 구한 다음 색상에 대해 주석을 기술하면 적절히 검색을 수행하는 시스템이다.

QVE[11]은 Hirata와 Kato가 만든 내용 기반 화상 검색 시스템으로 화상에서 추출한 외곽선 데이터를 그대로 특징 데이터로 사용하여 검색에 이용한다.

그런데, 이러한 내용 기반 검색 시스템인 QBIC, Chaobt, QVE 등은 아주 제한적인 범위의 검색에는 효과적이지만 동영상의 특성상 방대한 양의 데이터가 담고 있는 일반적인 의미를 추출하지 못하기 때문에 현실적이지 못하다.

따라서 이러한 두 종류의 기법 즉, 주석 기반 검색과 내용 기반 검색을 통합적으로 사용할 수 있는 방안의 모색이 필요하다. 이때 사용자의 요구에 적합한 형태로 두 가지 검색 기법을 통합적으로 지원하기 위해서는 무엇보다 통합 데이터 모델의 개발이 선행되어야 한다.

최근, 이 두 기법들을 통합하는 형태의 검색을 위해 김기병 등[2]은 다중 계층 동영상 모델(Multi-Layered Video Model; MuVi)을 제안하였다. MuVi 모델은 동영상 스트림 계층, 특징 계층, 특징 시퀀스 계층, 부개념 계층, 개념 계층으로 나누어져 각 계층별로 독립성을 유지하여 내용 기반이나 주석 기반 검색의 특정 방법에 종속되지 않는 질의 처리기를 구현하였다. 그러나 동영상 데이터 자체와 부여된 메타 정보 및 추출된 특징 정보들이 혼합되어 일직선상의 계층을 이루고 있으며, 개념 계층인 주석 기반 검색을 위한 모델링이 미흡하다.

본 논문에서는 동영상의 물리적인 구조에서 의미있는 개념 계층으로 분할하는 MuVi 모델과는 달리 동영상의 논리적인 데이터를 기준으로 사용자의 요구를 넓은 범위에서 좁은 범위로 줄여 갈 수 있도록 상위 수준에서는 여러 단계를 거쳐 주석을 부여하도록 분할하고(주석 기반 검색에 이용), 하위 수준에서는 동영상 데이터의 특징을 추출해 낼 수 있도록 분할하는 방법(내용 기반 검색에 이용)을 사용한다.

표 1은 본 논문에서 제안하는 IVDM 모델과 가장 유사한 MuVi 모델을 비교 분석한 것이다.

3. 통합 동영상 데이터 모델(IVDM)

동영상을 효율적으로 검색하고 관리하기 위해서

표 1. MuVi 모델과 IVDM 모델 비교 분석

	MuVi	IVDM
지원하는 검색기법	주석기반검색, 내용기반검색, 통합 검색	주석기반검색, 내용기반검색, 통합 검색
통합검색의 접근 방법	내용기반을 주석기반에 직접 대응시킬 때 발생하는 의미상의 차이를 해결하기 위해 부개념 계층을 도입	검색 범위를 넓은 범위에서 좁은 범위로 줄여 감 (상위 수준 : 주석기반 검색, 하위 수준 : 내용 기반 검색)
동영상 분할과정	추상화 과정 (물리적 데이터 → 의미 데이타)	구체화 과정 (논리적 데이터 → 물리적 데이타)
질의 유형	5가지	4가지
장점	계층별 독립성, 특징 벡터와 개념에 의한 내용기반 검색 가능	질의 유형별 독립성, IVDM만으로 데이터베이스스키마 구조 생성 가능
보완점	개념계층과 자료구조에 대한 보다 자세한 기술이 필요함	주관적인 주석을 배제하기 위한 정교한 프레임워크가 필요함

는 압축된 동영상 자체, 그에 관련된 주석, 그리고 화상 분석 결과를 통합적으로 데이터베이스화하여 공유할 필요가 있다. 이를 위해서는 일반적인 표준 모델을 만들어 다양하고 방대한 양의 압축 동영상을 관리하는 작업이 요구된다. 따라서 이 장에서는 동영상 정보의 관리를 위한 통합 동영상 데이터 모델(Integrated Video Data Model, IVDM)을 제안한다.

이러한 IVDM의 의의는 동영상 데이터를 구조화함으로써 다양한 동영상 데이터에 대해 상위 수준에서는 자유로운 주석 기반 검색을 지원하고 하위 수준에서는 내용 기반의 검색을 지원한다는 것이다. 그리고 이 두 가지의 단점을 서로 보완하고 장점을 극대화하기 위해 통합 형태의 검색도 동시에 지원한다. 이러한 작업으로 동영상 정보를 공유하여 사용할 수 있게 된다.

3.1 동영상의 분석

이 절에서는 특정 동영상 데이터만을 관리하는 것이 아니라 가능하면 많은 종류의 동영상 데이터를 관리할 수 있는 동영상 데이터 모델을 제시하고자 한다. 이를 위해서는 전형적인 동영상 데이터를 분석하여, 일반적인 동영상 데이터 모델을 이끌어 내는 방법을 사용한다. 따라서, 여기서는 많은 동영상 데이터들 중에서도 가장 쉽게 접할 수 있고 동영상 데이터의 성격을 가장 전형적으로 갖고 있는 뉴스를 예로 주석 기반 및 내용 기반 검색에 대한 요구를 분석한다.

동영상에 대한 주석 기반 검색을 위해서는 주석을 부여하는 단위가 중요하다. 이의 결정에 영향을 미치는 요소는 크게 두 가지이다. 첫째는 동영상에 대한 자동적인 내용 분석이 적용되기 힘든 수준이며, 둘째는 주석 부여자가 주석을 달기 용이한 수준이다. 따라서 주석 기반 검색은 동영상을 분할하는 단위의 상위 수준에 적용하는 것이 효과적이다.

동영상 분할 과정의 최상위 수준에 해당되는 것으로 동영상을 주제별로 분류할 수 있다. 이때 각 주제별 메타 데이터를 주석으로 부여하여 검색에 이용할 수 있다.

뉴스를 예로 들면, 1회 방송 분량을 하나의 단위로 보고, 최상위 수준에서 뉴스 전체에 방송 날짜, 방송 시각, 앵커 이름 등의 주석을 부여한다. 다음 단계로 뉴스를 경제, 사회, 정치, 스포츠 등의 주제별로 나눈다. 그리고, 각 주제별로 세부적인 사건들로 나눌 수 있고, 하나의 사건도 여러 명의 기자별로 나눌 수 있다. 만약 주제가 사회 분야라면 수해 상황이나 IMF 영향에 따른 실직자 증가 등이 사건이 될 수 있다. 그리고 수해 상황이라는 하나의 사건에 대해 여러 명의 기자가 지역별로 취재를 할 수 있다.

여기는 단지 뉴스를 예로 들었지만 다른 종류의 동영상이라면 나누는 단위나 단계가 다를 것이고, 각 단위별로 부여되는 주석 내용도 달라질 것이다. 그러므로 종류별로 고유의 특성을 잘 반영하는 표준을 세워둘 필요가 있다.

주석 기반 검색이 동영상의 추상 수준의 검색이라면 내용 기반 검색은 동영상을 구성하는 각각의 프레임을 분석한 결과를 이용하는 것이기 때문에 물리적인 수준 즉, 동영상 분할 과정의 하위 수준에 해당한다.

내용 기반 검색은 크게 2가지 차원 즉, 시간 차원과 공간 차원으로 분류될 수 있다[10]. 뉴스를 예로 앞의 주석 기반에 이어 내용 기반 검색을 위해 분석해 보면, 한 사건에 대한 기자의 취재 내용은 단순한 사건 장면을 보여주는 것인지 인터뷰 장면인지, 기자가 직접 나타나 설명을 하는 부분인지 등에 따라서 나눌 수 있다. 이 나누어진 부분에서도 중심 객체의 움직임에 관심이 있을 수도 있고, 아니면 특정 프레임에 등장하는 객체 자체에 관심이 있을 수도 있다. 예를 들면, 박세리 선수의 우승을 취재한 기자가 있다면 그 취재 내용 중에는 박세리 선수가 필드 위를 걷는 모습이나 스윙하는 모습, 인터뷰하는 모습 등을 담은 동영상이 있을 수 있다. 이 중 스윙하는 내용에서는 여러 개의 객체 즉, 박세리 선수, 골프공, 골프채 등이 나타난다. 여기에서 박세리 선수의 스윙이나 골프공이 날아가는 모습에 관심이 있는 경우는 시간 차원의 질의에 해당할 것이고, 박세리 선수의 옷차림이나 골프채 모양 등에 관심이 있는 경우는 공간 차원의 질의에 해당될 것이다.

그러므로 내용 기반의 검색을 위해서는 시간 차원의 특징인 움직임 특징 데이터와 공간 차원의 특징인 색깔, 모양, 질감 특징 데이터가 필요하다. 이 각각의 특징을 분석해 내는 방법은 본 논문의 범위를 벗어나기 때문에[5,9,10] 본 논문에서는 생략한다.

3.2 IVDM의 구조

<그림 1>은 일반적인 동영상을 분할하는 과정을 나타내고 있다. 동영상 분석을 위해서는 우선 이러한 분할 과정이 요구된다. 우리의 분할 과정은 동영상을 영화, 뉴스, 드라마, 화상회의 등의 종류별로 분류했을 때 비디오 스트림이 이들 중 하나의 종류에 속한다는 전제하에 이루어진다. 뉴스 전체는 비디오스트림(Video_Stream)이 되고 주제별 나눔이나 사건, 기자별 나눔은 주제 단위(Thematic_Unit)가 되는데, 왼쪽에 있는 순환 화살표가 여러 단계까지 반복적으로 나누어질 수 있다는 의미이다. 이때 반복 회수는 동영상의 종류에 따라 달라질 수 있고, 주제 단위의 개수나 크기도 동영상의 종류나 주제마다 달라질 수 있다. <그림 1>에서 가장 작은 단위의 주제 단위를 시퀀스(Sequence)라고 표현한다. 앞의 예와 연관지으면 시퀀스는 한 기자의 취재 내용이 된다. 시퀀스를 다시 씬(Scene)으로 나누는데, 여기서 씬은 앞서의 한 기자의 취재 내용을 단순 사건 장면인지 혹은 인터뷰 장면인지 등에 따라 나누어진 부분에 해당된다. 동영상의 흐름 순서 즉, 시간 차원의 검색을 위해 이 씬에서 일정한 간격으로 프레임을 추출하여 SI_프레임(Same Interval frame)이라 칭한다. 공간 차원의 검색을 위해서는 각 씬에서 대상 객체가 존재하는 구간 즉, 세그먼트(Segment)로 나누고 이 중에서 대상 객체가 가장 뚜렷이 나타나는 프레임을 키-프레임으로 삼는다. 추출된 SI_프레임에서는 카메라의 움직임이나 객체의 움직임을 분석해 내고, 키-프레임에서는 색깔, 모양, 질감 등을 분석해 내어 검색시 이용한다.

<그림 1>에 나타나는 과정을 통해, 우리는 <그림

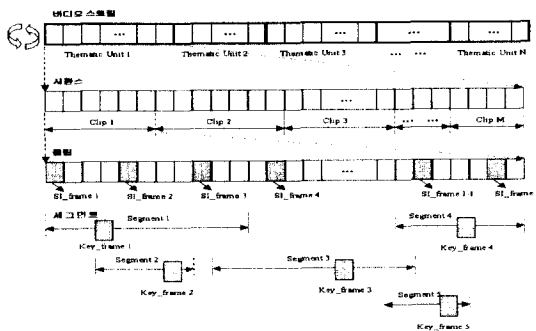



그림 1. 동영상의 특징 추출을 위한 분할 과정

2>와 같은 통합 동영상 데이터 모델(IVDM)을 제안한다. <그림 2>는 OMT 객체도[4]로 기술되었다. OMT 객체도는 클래스와 그들 사이의 관계를 나타내며 이는 데이터베이스의 설계에도 매우 적합한 것으로 알려져 있다[14].

3.2.1 IVDM 클래스들의 의미

IVDM 클래스는 모두 22개이다.

1) Video_Document 클래스 : 원래의 동영상 데이터 자체(raw video)이다.
2) Video_Stream 클래스 : Video Document를 Video Stream으로 표현한다.
3) Stored_VideoSegment 클래스 : 대용량의 Video Stream을 나누어 검색에 용이한 형태로 DB에 저장하는 단위이다.
4) Thematic_Unit 클래스 : 검색을 위해 주석을 부여하는 단위로서 동영상의 주제별로 구분한다. 1차로 주제별로 나누어진 동영상을 더 작은 단위의 부주제로 나눌 필요가 있을 때는 이미 나누어진 Thematic_ Unit을 다시 더 작은 Thematic_Unit으로 반복해서 나눌 수 있다. 이때 동영상의 종류에 따라 Thematic_ Unit 단계가 다양하게 반복될 수 있고, 단위 크기도 달라질 수 있다. 만약 동영상의 종류가 영화라면 시나리오 상에서 Scene이나 Shot, Cut 등으로 구분되는 것까지 이 단계에 포함시킬 수 있다. 여기서 Scene은 큰 Thematic_Unit이 되고 더 작은 Shot이라는 1개 이상의 Thematic_Unit으로 구성된다. 그리고, 각각의 Shot은 1개 이상의 Cut이라는 아주 작은 Thematic_Unit으로 이루어진다.
5) Annotation 클래스 : Video_Stram과 Thematic_Unit에 설명을 부여하고자하는 주석의 타입과 이름을 입력하는 클래스이다. 고정된 타입이나 이름의 주석이 아니라 데이터베이스 스키마를 구성할 때 동영상의 성격에 맞게 유동적으로 부여한다.
6) Sequence 클래스 : 나누어진 가장 작은 단위의 Thematic_Unit으로 연속적인 프레임의 집합이다.
7) Scene 클래스 : 같은 장소와 같은 시간대를 배경으로 하는 구간이다.

8) SI_frame 클래스 : 프레임 Sequence에서 일정한 간격으로 추출한 프레임으로 동영상의 움직임을 분석하기 위한 프레임이다.
9) Type 클래스 : 객체의 움직임 형태를 동사 원형으로 표현해두고, 그 속성은 boolean형으로 하여 대상 객체가 속하는 타입에 yes로 표시한다.
10) T_feature 클래스 : Camera_Motion과 Object_Motion 클래스로 일반화된다.
11) T_keyword 클래스 : 현재의 SI_frame과 바로 직전의 SI_frame을 편집자가 직접 비교하여 간단한 keyword로 객체의 움직임 형태를 입력한다.
12) Camera_Motion 클래스 : 현재의 SI_frame과 바로 직전의 SI_frame에서 움직임 분석 알고리즘을 이용하여 카메라의 움직임 변화를 추출하여 결과를 저장하는 클래스이다.
13) Object_Motion 클래스 : 현재의 SI_frame과 바로 직전의 SI_frame에서 움직임 분석 알고리즘을 이용하여 객체의 움직임 변화를 추출하여 결과를 저장하는 클래스이다.
14) Segment 클래스 : Scene에서 동일한 객체가 존재하는 구간이다.
15) Key_frame 클래스 : Segment에서 가장 객체가 뚜렷이 드러나는 프레임이다.
16) S_keyword 클래스 : key_frame 이미지를 대표할 수 있는 내용을 간단한 keyword로 나타낸다.
17) Category 클래스 : 모양, 색깔, 질감을 text 기반으로 검색 가능하도록 하기 위한 클래스이다.
18) C_shape 클래스 : key_frame에 나타나는 객체의 모양을 보고 객체의 종류와 구체적인 이름을 저장하는 클래스이다.
19) C_color 클래스 : 모양이 지정된 객체의 색깔을 표현한다.
20) C_texture 클래스 : 모양이 지정된 객체의 질감을 표현한다.
21) S_feature 클래스 : Color, Texture, Shape 클래스로 일반화된다.
22) Color, Texture, Shape 클래스 : Key_frame에서 화상 분석 알고리즘을 이용하여 화상의 특징 즉, 색깔, 질감, 모양 벡터를 추출해 내어 그 결과를 저장하는 클래스이다.

3.2.2 IVDM 클래스들 사이의 관계

IVDM 클래스들 사이의 관계는 <그림 2>에 잘 나타나 있다. <그림 2>에서 제일 위의 Video_Document는 Video_Stream으로 다시 1:1 표현되며 이는 1개 이상의 Stored_VideoStream으로 구성되어 데이타베이스에 저장된다. 이 때 Video_Stream은 시작 프레임과 끝 프레임을 나타내는 두 가지 애트리뷰트를 가진다. Video_Steam과 Annotation 및 Thematic_Unit 사이는 부품 관계(part_of)로서, 1개 이상의 Annotation과 Thematic_Unit의 집합이 Video_Stream이 된다. Thematic_Unit은 다시 더 작은 Thematic_Unit을 포함할 수도 있고, 아닐 수도 있으며 Video_Stream과 마찬가지로 1개 이상의 Annotation으로 구성된다. 또 이 Thematic_Unit을 Sequence로 다시 표현하고, Sequence는 0개 이상의 Scene으로 이루어진다. 0개 이상이므로 Sequence 이하에 Scene이 없을 수도 있는데, 이것은 주석 기반의 검색만 가능하게 되는 것이고 적어도 하나의 Scene이라도 존재한다면 이는 내용 기반 검색까지 지원하게 되는 것이다. Scene과 Segment 및 SI_frame 사이는 일반화

그림 2. IVDM의 구조

관계(is_a)로서, Scene을 Segment나 SI_frame으로 다시 표현할 수 있다. SI_frame은 각각 Type, T_feature, T_keyword로 표현할 수 있고, T_feature는 다시 Camera_Motion과 Object_Motion으로 일반화된다. Segment는 Key_frame과 참조 관계(association)에 있고, 이 Key_frame은 다시 Category, S_keyword, S_feature로 표현할 수 있다. 이중 Category는 0개 이상의 C_shape를 가지고, 이는 각각 0개 이상의 C_color와 C_texture로 구성된다. 마지막으로 S_feature는 Color, Shape, Texture로 일반화된다.

4. 뉴스 동영상의 스키마 설계

앞에서 제시한 IVDM 모델을 기반으로 뉴스 동영상의 스키마 구조와 이에 알맞은 질의 유형을 설계하고, 그 처리 과정을 알아본다. 실제 구현은 인덱스 정보 관리를 위해 객체관계형 DBMS인 Informix 9.*를 사용하였고, Visual C++ 5.0을 사용하여 사용자 인터페이스 부분을 구현하였으며, DBMS와의 연결은 ODBC를 사용하였다.

4.1 뉴스 동영상의 스키마 구조

이 절에서는 3장에서 제시한 IVDM 모델을 바탕으로 동영상의 대표적인 예가 될 수 있는 뉴스 동영상을 가지고 객체 지향 데이타베이스에서 구현 가능하도록 설계한다.

뉴스 동영상의 인덱스 관리를 위해 객체 지향 개념을 도입한 이유는 다음과 같은 장점이 있기 때문이다. 첫째, 메타데이타 구조를 사용하여 비슷한 특징을 가진 객체의 클러스터링(clustering)을 증진시킨다. 둘째, 이 스키마를 사용하는 질의는 특정 클래스나 클래스 계층으로 검색 공간을 제한하게 되므로 검색 성능이 향상되고, 더 정확한 검색 결과를 얻을 수 있다. 셋째, 상속 관계로 분류되므로 공통 특성을 가진 객체를 관리하기가 편리하다. 넷째, 스키마 구조에 관한 사전 지식이 없는 초보자들이 검색하기 쉽다.

이 스키마 구조 역시 OMT 객체도의 형태로 기술되었다. 각 네모 상자의 윗 부분은 데이타베이스의 클래스명이고, 아래 부분은 각 클래스의 속성들이다.

뉴스 동영상 스키마는 제일 상위의 video 클래스의 속성인 시작 프레임, 끝 프레임 및 실제 비디오 데이타의 oid를 각 하부 클래스가 상속받는다. 그리고, news, theme, event, reporter, scene 클래스는 주석 기반 검색을 위한 클래스로서 순서대로 oid를 갖는 속성으로 연결되어진다. scene 이하 클래스는 내용 기반 검색을 위한 클래스로 key_frame 이하는 공간 차원의 검색을 위한 것이고, SI_frame 이하는 시간 차원의 검색을 위한 클래스들이다. 특히, category 클래스는 키프레임의 색깔, 모양, 질감을 texture 형태로 표현하여 검색하기 위한 것으로 먼저 객체의 모양을 먼저 파악한 후 동물, 식물, 자연, 빌딩 등의 큰 종류로 분류한 후 그 객체의 색깔이나 질감은 oid로 연결한다. 물론 질감이나 색깔로도 바로 검색 가능하다. 실제 이미지 분석 알고리즘을 통한 자동화된 방법을 이용하기 위한 클래스는 s_feature나 t_feature 이하의 클래스가 된다.

4.2 IDVM의 질의 유형

이 절에서는 앞에서 제시한 IVDM 모델에 근거해 뉴스 동영상 검색에 사용할 수 있는 질의 유형을 제안하고, 질의 처리 과정을 보여준다.

그림 3. 뉴스 동영상의 스키마 구조

4.2.1 질의의 종류

IVDM 모델에 기반한 단계별 뉴스 동영상의 질의 유형은 4가지로 분류된다.

1) Type-1 : Thematic_Unit 기반의 질의(주석 차원)
큰 주제에서 작은 주제로 단계별로 부여된 주석을 기반으로 검색한다.

예1) "1998년 7월" "경기 지역" "수해" 상황을 검색하라
SELECT video, s_frame, f_frame FROM event WHERE (when_date >= '1998/07/01' AND when_date < '1998/08/01') AND (where IN '경기도') AND what IN '수해';

2) Type-2 : SI_frame 기반의 질의(시간 차원)
시간의 흐름에 따른 객체의 움직임에 대해 검색한다. 검색하고자 하는 객체의 움직임 형태를 키워드나 동작을 나타내는 동사 원형 형태로 질의할 수도 있고, 실제 그 장면이 나타나는 동영상을 질의로 줄 수도 있다.

예2) "뛰는" 모습이 담긴 동영상을 검색하라(motion_type 기반).
SELECT video, s_frame, f_frame FROM motion_type WHERE run = TRUE;

예3) 뛰는 모습이 담긴 동영상을 검색하라(분석 알고리즘을 이용).
SELECT video, s_frame, f_frame FROM obejct_motion WHERE |QUERY:동영상 이동벡터 - DB:동영상 이동벡터| < 임계치;

3) Type-3 : Key_frame 기반의 질의(공간 차원)
특정 객체가 나타나는 장면을 검색한다. 화상에 나타나는 객체의 타입과 이름, 색깔, 모양, 질감 등을 키워드나 카테고리 형태로 질의할 수도 있고, 직접 그림을 그리는 툴을 이용하여 스케치한 화상이나 파일의 형태로 존재하는 화상을 가져와서 검색할 수도 있다.

예4) "신창원"의 몽타주가 나타나는 장면 검색을 검색하라(s_keyword 기반).
SELECT video, s_frame, f_frame FROM s_keyword WHERE (keyword IN '신창원') AND (keyword IN '몽타주');

예5) 신창원의 사진 화상으로 검색으로 검색하라 (분석 알고리즘 이용).
SELECT video, s_frame, f_frame FROM color, texture, shape WHERE |QUERY:화상의 특징값 - DB:화상의 특징값| < 임계치;

4) Type-4 : 통합된 질의
앞의 3가지 타입을 통합한 형태의 질의이다. 이 통합 형태는 하나의 타입형 질의로 검색한 결과를 도메인으로 다른 타입의 검색을 하므로 단계별로 검색 범위를 줄여 주게되어 더 빠르고 정확한 검색이 가능하다.

예6) "1998년 8월"에 "고종수" 선수가 "뛰는" 모습이 담긴 뉴스를 검색하라(Type-1과 Type-2의 결합).
Ref1 := SELECT * FROM event, t_keyword, type WHERE (when_date>='8/1/1998') AND (when_date<'9/1/1998') AND (who IN '고종수');
Ref2 := SELECT * FROM type WHERE run = TRUE;
(Ref1의 범위 <= Ref2의 범위)인 조건만 검색 결과로 함;

예7) "박세리" 그리고 골프를 치는 사람이 있는 화상으로 검색하라(Type-1과 Type-3의 결합).
Ref1 := SELECT * FROM event WHERE who IN '박세리';
Ref2 := SELECT * FROM color, texture, shape WHERE |QUERY:화상의 특징값 - DB:화상의 특징값| < 임계치;
(Ref1의 범위 <= Ref2의 범위)인 조건만 검색 결과로 함;

4.2.2 질의 처리 알고리즘

사용자의 질의를 처리하는 과정을 개괄적인 알고리즘으로 나타내면 다음과 같다.

[알고리즘 1] 사용자 질의 처리 알고리즘
CASE Type OF {질의의 유형에 따라 4가지로 분

```
기)
  Type-1 :{질의 유형이 주석기반 질의이면,}
    Result = Search(Q);{단순 SQL문으로 검색}
  Type-2 :{질의 유형이 SI_frame 기반의 질의이
        고,}
    IF Keyword 방식 {키워드형 질의이면,}
      THEN Result = Search(Q);{SQL문을 통해
            DB의 자료를 검색}
    ELSE BEGIN{키워드 방식이 아니면,}
      Result = Analysis(Q_Video);{질의 동영상을
            분석}
      Result = Search(Result);{결과를 DB에 있는
            내용과 비교 검색}
    END
  Type-3 :{질의 유형이 Key_frame 기반 질의이
고,}
    IF Keyword 방식{키워드형 질의이면,}
      THEN Result = Search(Q);{SQL문을 통해
            DB의 자료를 검색}
    ELSE BEGIN{키워드 방식이 아니면}
      Result = Analysis(Q_Image);{질의 화상 분
            석}
      Result = Search(Result);{분석한 결과를 DB
            이 있는 내용과 비교 검색}
    END
  Type-4 :{질의 유형이 혼합된 형태이면}
    BEGIN
      Result1 = Search_Type1(Q);{Type1 질의를
            검색}
      Result2 = Search_Type2(Q);{Type2 질의를
            검색}
      Result3 = Search_Type3(Q);{Type3 질의를
            검색}
      Result = Compare(Result1, Result2);{Type1
            과 Type2의 질의 결과를 비교하여
            교집합을 구함}
      Result = Compare(Result, Result3);{바로
            위의 검색결과와 Type3의 질의 결
            과를 비교하여 교집합을 구해 최종
            결과를 냄}
    END
  OTHERWISE {위의 4가지 유형의 질의가 아니면}
    Error{에러 처리}
  END.

  FUNCTION Compare(temp1, temp2)
  BEGIN {두 가지 유형의 검색 결과를 비교하여 교
집합을 구하는 함수이다}
    WHILE(temp1 != null) BEGIN
      WHILE(temp2 != null) BEGIN
        IF(temp1.video == temp2.video)
        THEN IF((temp1.s_frame<=temp2.s_frame)
              AND (temp1.f_frame>=temp2.
              f_frame))
            THEN Result = temp1;
        temp2++;
      END
      temp1++;
    END
    RETURN Result;
  END.
```

4.3 뉴스 동영상을 검색하는 과정

지금까지 설계한 내용을 기초로 뉴스 동영상 검색의 실제 처리 과정은 <그림 4>와 같다. <그림 4>는 MPEG-2로 압축된 뉴스 동영상을 대상으로 한 경우로 크게 사용자 인터페이스, 동영상 처리 모듈, 데이

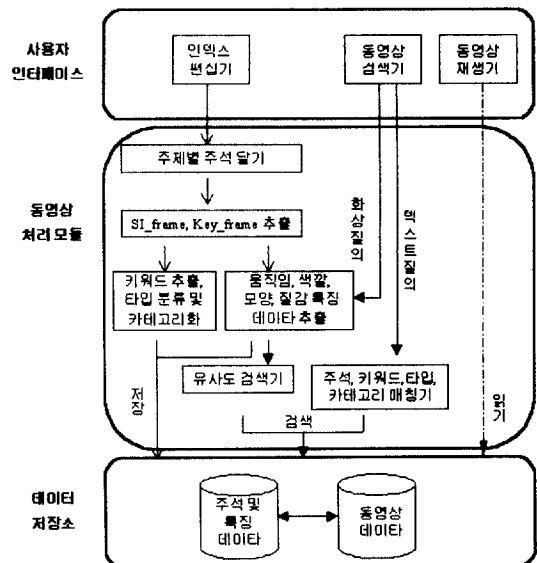



그림 4. 뉴스 동영상 검색 과정

터 저장소로 구성된다. 사용자 인터페이스는 다시 인덱스 편집기, 동영상 검색기 및 동영상 재생기로 이루어진다.

사용자 인터페이스에서 사용자의 요구에 응답하기 위해 데이터 저장소인 DBMS에 있는 실제 동영상 데이터나 관련 정보에 접근하기까지의 과정이 동영상 처리 모듈이다. 먼저, 인덱스 편집기의 역할은 나중의 주석과 I-프레임에서 내용 기반 동영상 검색을 위해 주제별로 주석을 달아두고, SI_frame과 Key_frame을 추출해 낸다. SI_frame이나 Key_frame의 후보 키가 되는 것은 특성상 MPEG-2 압축 동영상을 구성하는 I-, B-, P-프레임 중에서 I-프레임만 후보 프레임이 될 수 있다. 추출해낸 SI_frame과 Key_frame을 분석하여 키워드 형태의 특징 타입과 움직임에서는 움직임 타입을, 색깔, 모양, 질감 특징으로는 카테고리 형태로 관련 정보를 부여한다. 그리고, 실제로 움직임, 색깔, 질감 모양 특징 데이터는 동영상 분석 알고리즘을 통해 계산해내어 데이터 저장소의 주석 및 특징 데이타를 데이타베이스에 저장한다. 이렇게 저장된 내용들을 가지고 동영상 검색기에서 검색이 이루어진다.

동영상 검색기에서는 검색을 위해 입력된 데이터가 텍스트 형태인지 화상이나 동영상인지에 따라 검색 방법에 차이가 있다. 검색어가 텍스트 형태인 경우는 주제별로 부여된 주석 데이터나 내용 기반의 검색 중에서도 키워드, 움직임 타입, 카테고리 형태로 부여된 데이터를 대상으로 정확히 단어가 일치될 때 검색된다. 반면 화상이나 동영상 데이터가 질의로 입력되었을 경우에는 내용 기반 검색을 위해 프레임을 분석할 때처럼 질의 화상에서도 색깔, 모양, 질감, 움직임 등을 분석해서 데이타베이스에 저장되어 있는 특징 데이터와 비교한다. 이 때는 텍스트 형태의 질의와 같은 정확한 일치는 기대할 수 없고, 두 데이터의 차이를 구하여 지정한 임계치를 벗어나지 않는 범위의 동영상을 검색 결과로 보여준다.

검색된 결과나 인덱스 편집시에 동영상 재생기를 통하여 동영상을 보게 되는데, 이때는 단순히 데이타베이스에 저장된 동영상 데이터를 읽어와서 재생한다.

다음은 실제 질의와 그 처리 과정 및 검색 결과를 순서대로 나타낸다.

▶ 질의 : 1998년 12월에 일어난 스포츠계 사건들 중에 박세리 선수가 스윙하는 모습과 박찬호 선수가 투구하는 모습에 관한 내용을 검색하라.

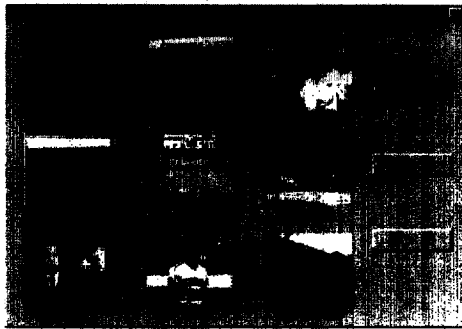
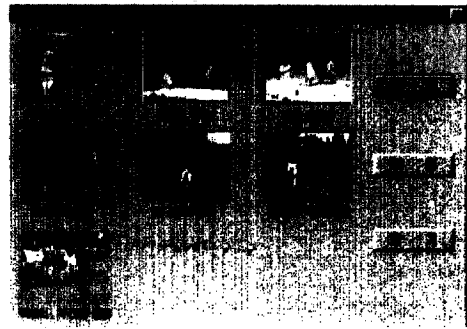
▶ 질의 처리 과정 :

```
Ref1 := SELECT * FROM  theme
        WHERE (when_date>='12/1/1998') AND
        (when_date<='12/31/1998')
          AND (kind = '스포츠');
Ref2 := SELECT * FROM  motion_type
        WHERE (swing = True) OR (throw =
        True);
Ref3 := SELECT * FROM  c_shape
        WHERE (name IN '박찬호') OR (name
        IN '박세리');
Temp := Compare(Ref1, Ref2);
Result := Compare(Temp, Ref3);
```

▶ <그림 5a>는 Type-1형(주석 기반) 질의 결과 즉, Ref1의 내용이고, <그림 5b>는 Type-2형(SI_frame 기반) 질의 결과인 Ref2의 내용이다. <그림 5c>는 Type-3형(Key-frame 기반) 질의 결과 즉, Ref3의 내용이다. 그리고 <그림 5d>는 앞서의 세 가지를 통합한 Type-4형 질의의 최종 결과인 Result의 내용이다.

5. 결론 및 향후 연구 과제

최근 방대한 양의 멀티미디어 정보가 쏟아져 나오고 있는데, 그 중에서도 동영상 정보가 큰 비중을 차지한다. 따라서 동영상 정보에 대한 검색 요구가 많아지고 있다.

본 논문에서는 이러한 동영상 정보를 효율적으로 검색하기 위해, 이전 연구들의 2가지 접근법 즉, 주석 기반 검색과 내용 기반 검색의 장점을 모두 수용하는 통합 동영상 데이터 모델(IVDM)을 제시하였다. IVDM은 동영상의 논리적인 데이터를 기준으로 사용자의 요구를 넓은 범위에서 좁은 범위로 줄여 갈 수 있도록 상위 수준에서는 여러 단계를 거쳐 주석을 부여하도록 분할하고(주석 기반 검색에 이용), 하위 수준에서는 동영상 데이터의 특징을 추출해 낼 수 있도록 분할하는 방법(내용 기반 검색에 이용)을 사

(a) Type-1형

(b) Type-2형

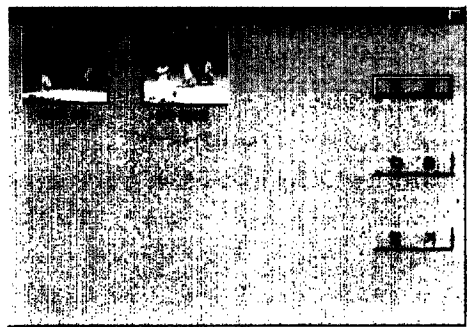
(c) Type-3형

(d) Type-4형

그림 5. 동영상에 대한 질의 결과

용한다. 이 모델을 기반으로 뉴스 동영상을 예로 객체 지향 데이타베스 스키마를 설계하여 이제 적합한 효율적인 검색 질의 형태 4가지를 제안하고, 질의 처리 과정을 알리고리즘으로 나타내었으며, 이를 토대로 실제 구현할 경우 전체적인 작업 과정을 나타내었다.

이러한 연구 결과는 주석 기반과 내용 기반의 장점을 최대한 수용할 수 있는 통합 모델을 제시하였다는 점과 이 모델을 검색 속도와 효율성을 고려해 객체 지향 데이타베이스 스키마로 설계하였다는 점에서 의의가 있다.

한편 본 논문에서는 인덱스 정보만 데이터베이스화해서 관리하는데 치중하였는데, 앞으로 동영상 데이터 자체도 데이타베이스에 구조화하여 저장할 수 있도록 확장되어야 한다. 그리고, 내용 기반 검색을 위한 씬이하 구조 중 키프레임에서의 카테고리와 SI_프레임에서의 타입을 표준화하기 위한 노력이 필요할 것이다.

참 고 문 헌

[1] 허진용, 김인홍, 배종민, 강현석, "MPEG-2 압축 동영상 정보 관리 시스템의 설계 및 구현", 정보처리학회 논문지, 제5권, 제6호, pp. 1431-1440, June 1998.

[2] 김기병, 김형주, "내용 기반 검색 및 주석 기반 검색을 통합하는 비디오 데이터 모델의 설계 및 구현," 정보과학회 논문지(C), 제3권 제2호, pp. 115-126, Apr. 1997.

[3] 김기욱, 김형주, "비디오 주석 시스템의 설계 및 구현," 정보과학회 논문지, 제24권 제6호, pp. 588-597, June 1997.

[4] J. Rumbaugh, M. Blaha, W. Premerlani, F. Eddy, and W. Lorensen, "Object-Oriented Modeling and Design," Prentice-Hall, 1991.

[5] R. and J. Gray, "Similar-shape Retrieval in Shape Data Management," IEEE Computer, pp. 57-62, Sept. 1995.

[6] J. Smith and S. Chang, "VisaulSEEK: a Fully Automated Content-based Image Query System," ACM Multimedia '96, Nov. 1996.

[7] A. Ono, M. Amano, M. Hakaridani, T. Satou, and M. Sakauchi, "A Flexible Content-Based Image Retrieval System with Combined Scene Description Keyword," IEEE, pp. 201-208, June 1996.

[8] Y. Kim and M. Shibata, "Content-Based Video Indexing and Retrieval - A Natural Language Approach," IEICE Trans. Inf. & Syst., E79-D (6), pp. 695-705, June 1996.

[9] V. Ogle and M. Stonebraker, "Chabot: Retrieval From a Relational Database of Images," IEEE Computer, 28(9), pp. 40-48, 1995.

[10] W. Niblack, R. Barber, W. Flickner, E. Glasman, D. Petkovic, C. Faloutsos, and G. Taubin, "The QBIC Project: Querying Image by Content Using Color, Texture, and Shape," Proc. SPIE Storage and Retrival for Image and Video Database, pp. 173-187, Feb. 1993.

[11] K. Hirata and T. Kato, "Query by Visual Example Content-based Image Retrieval," Advances in Database Technology(EDBT '92), pp. 56-71, 1992.

[12] B. Bruegge, J. Blythe, J. Jackson and J. Shufelt, "Object-Oriented System Modeling with OMT," OOPSLA '92, 27(10), PP. 414-427, Oct. 1992.

전 미 경

1996년 경상대학교 컴퓨터과학과 이학사
1999년 경상대학교 전자계산학과 공학석사
관심분야 : 객체지향 데이터베이스, 동영상 정보 검색

김 인 홍

1983년 경북대학교 공과대학 전자공학과 공학사
1984년~1989년 한국전자통신연구원 연구원
1985년 한국과학기술원 전산학 석사
1989년~1990년 삼성종합기술원 선임연구원
1990년~1995년 MasPar Computer Corp.(미국) 선임연구원
1995년~1997년 Jinnes Technologies. Inc.(미국) 이사
1997년~현재 Hycomm Research 대표
1997년~현재 경상대학교 전자계산학과 박사과정 재학중
관심분야 : 멀티미디어 전송 및 응용 S/W, ATM Network, 동영상 처리 및 저장 검색

류 시 국

1980년 경북대학교 전자공학과 공학사
1989년 영남대학교 전자공학과 공학석사
1997년 경상대학교 전자계산학과 공학박사
1980년~현재 경남정보대학 컴퓨터정보과 교수
관심분야 : 객체지향 데이타베이스, 멀티미디어, 병렬처리

전 용 기

1980년 경북대학교 컴퓨터공학과 학사
1982년 서울대학교 전산과학과 석사
1993년 서울대학교 전산과학과 박사
1982년~1985년 한국전자통신연구소 연구원
1995년~1996년 미국 캘리포니아 주립대(UCSC) 연구교수
1985년~현재 경상대학교 컴퓨터과학과 교수
관심분야 : 병렬 및 분산 처리, 운영체제, 프로그래밍 환경

강 현 석

1981년 동국대학교 전자계산학과 학사
1983년 서울대학교 계산통계학과 이학석사(전산학)
1989년 서울대학교 계산통계학과 이학박사(전산학)
1984년~1993년 전북대학교 전자계산학과 부교수
1993년~현재 경상대학교 컴퓨터과학과 교수
관심분야 : 객체지향 데이터베이스, 전자문서 관리, 멀티미디어